Advances in Modeling Visual Search and Target Discrimination Performance

G. Witus
Turing Associates, Inc.
1392 Honey Run Drive, Ann Arbor, MI 48103

R.E. Karlsen, G.R. Gerhart, and D.J. Gorsich
U. S. Army TARDEC
Warren, MI 48397

Abstract

This paper reports on advances in mathematical models of observer-ensemble performance in narrow-field-of-view visual search and target discrimination for ground vehicles in natural terrain. Three developments are presented. We show that the distribution of search time follows a lognormal distribution. We show that search outcome is the result of a race between scene parsing and target detection. Scene parsing guides and focuses search, but also leads to quitting without detection. Quitting is not simply the consequence of having exhausted the supply of suspect locations. We present a refined target signature metric and computational method that emulates human perceptual organization of the target into component regions. This is significant not only because it improves the ease-of-use and reduces subjective user input, but also because it is potentially applicable to thermal images. The refined metric provides reasonably accurate prediction of probability of detection for cued detection and uncued search experiments.

1 Introduction

This paper reports on advances in mathematical models of observer-ensemble performance in narrow-field-of-view visual search and target discrimination for ground vehicles in natural terrain. These results build on and extend visual target discrimination performance modeling and perception experiments previously reported [1].

Three developments are presented. Section 2 shows that the distribution of search time follows a lognormal distribution, in contrast to the commonly-used negative exponential model. Section 3 shows that search dynamics have the characteristics of a race process between target detection and quitting based on parsing the scene into recognized non-target objects or regions. Section 4 presents a refined signature metric and computational method that emulates pre-attentive perceptual organization of the target. Section 5 discusses the implications and limitations of these advances.

2 Distribution of Search Time

The basic theory is that search involves two complementary activities: (1) recognizing background objects, and (2) examining suspect target locations. Recognizing background regions guides and focuses search by eliminating areas from further consideration, indicating promising locations to search, and by guiding expectations of apparent target size. When the scene is fully parsed into recognized background objects (to the observer's satisfaction), search is terminated by quitting without reporting a detection.

Search begins with a scan of the image to begin parsing and detect highly obvious targets. As the scene is parsed, background objects and regions are recognized and rejected (e.g., the sky, distant hillsides, close and empty foreground fields). Suspicious locations, when noticed, are examined. Eye movement studies have found evidence of two states in visual search [2].

The observer's thresholds to examine a suspect location and to decide to designate a target begin high and are lowered over time. This helps ensure that low-confidence suspect locations do not delay or pre-empt inspection and detection at higher confidence locations.

This dynamic implies that the hazard rate, i.e., the rate of search termination as a function of time given that search is still in process, has a characteristic shape:

- It begins at zero.
- It increases as scene parsing guides and focuses search.
- It tails off to zero as the supply of suspect locations is exhausted.

The hazard function for the lognormal distribution has the shape characteristics required by the search theory. The hazard function of the negative exponential distribution is a constant. Neither do generalizations of the negative exponential distribution, e.g., the Weibull distribution and Gamma distributions, have the required hazard function shape. The lognormal model has been used as a model of human information processing in other contexts [3].

Table 1 tabulates several measures of the goodness of fit between the empirical search time distribution and model distributions, for the distribution of search times for 1151 images. Each empirical search time distribution consists of 23 points, i.e. the search times for 23 observers. Four model distributions are compared: lognormal, negative exponential, Weibull, and the sum of two different negative exponential (i.e., a negative exponential delay followed by a negative exponential search time). The measures of fit are the root-mean-square (RMS) error in the cumulative distribution, the proportion of variance (PV) in the empirical distribution explained by the estimate, the maximum Kolmogorov-Smirnov (K-S) statistic over all 1151 images, the average K-S statistic, and the linear regression slope and intercept. The K-S statistic is a measure of the difference between continuous distributions (whereas the chi-squared distribution is used for discrete distributions). The K-S statistic is the maximum of the absolute value of the difference between the two cumulative distribution functions. The lognormal model is a better match than the other model by all measures of performance.

	Lognormal	Neg. Exponential	Sum 2 Neg. Exp.	Weibull
RMS Error	0.068	0.085	0.079	0.084
PV Explained	0.940	0.888	0.917	0.908
Maximum K-S	0.276	0.393	0.357	0.389
Mean K-S	0.127	0.168	0.145	0.152
Slope	1.000	0.823	0.972	0.711
Intercept	0.007	0.125	0.012	0.050

Table 1: Comparison of Search Time Distribution Models

The results of applying the W test of normality of a distribution [4] at different significance levels to the log search time for each of the 1151 images are summarized in Table 2. The significance level is the probability of rejecting the hypothesis of normality when the data are drawn from a normal distribution. The results indicate that the hypothesis of normality is rejected for seven percent of the images, and is questionable for another eight percent.

Percent Rejection for Samples of a Normal Distribution	Percent Rejection for Image Search Time Distributions	Difference
50%	56.9%	6.9%
40%	51.6%	11.6%
30%	43.3%	13.3%
20%	34.5%	14.5%
10%	25.0%	15.0%
5%	17.5%	12.5%
1%	8.3%	7.3%

Table 2: Results of the W Test for Normality of Log Search Time Distributions

The lognormal distribution has two parameters: the mean log time and the standard deviation of log time. In contrast, the negative exponential distribution has only one parameter: the rate which is equal to one over the mean time. The proposed application of the lognormal model uses one constant value of the standard deviation of log search time for all images, set equal to the mean over all images standard deviation of log search time (0.83). Thus the proposed lognormal model requires one parameter for each image (as does the negative exponential model), plus one additional constant for all images.

3 Race Between Quitting and Target Detection

The proposed model asserts that two processes are active during search: recognition and rejection of background objects/regions (exclusive processing) and location and examination of suspect regions (inclusive processing). Individuals employ a mix of these two strategies. Observers relying more on inclusive processing will tend to quit only after exhausting the supply of suspect locations, and will tend to have longer quitting times than detection times. Observers relying more on exclusive processing will tend to take less time to come to the decision to quit than to detect difficult targets. Obvious targets may be detected during the scene parsing. In general, observers will employ a mix of these strategies. They may shift between strategies during search of a single scene, and may change emphasis from scene to scene.

Data from the search experiment supports the proposed model. Figure 1 shows that some subjects tend to have longer quitting times than detection times, while other subjects tend to have shorter quitting times than detection times. (The figure also shows that the search time for low-confidence detections is consistently longer than the search time for high-confidence detections.)

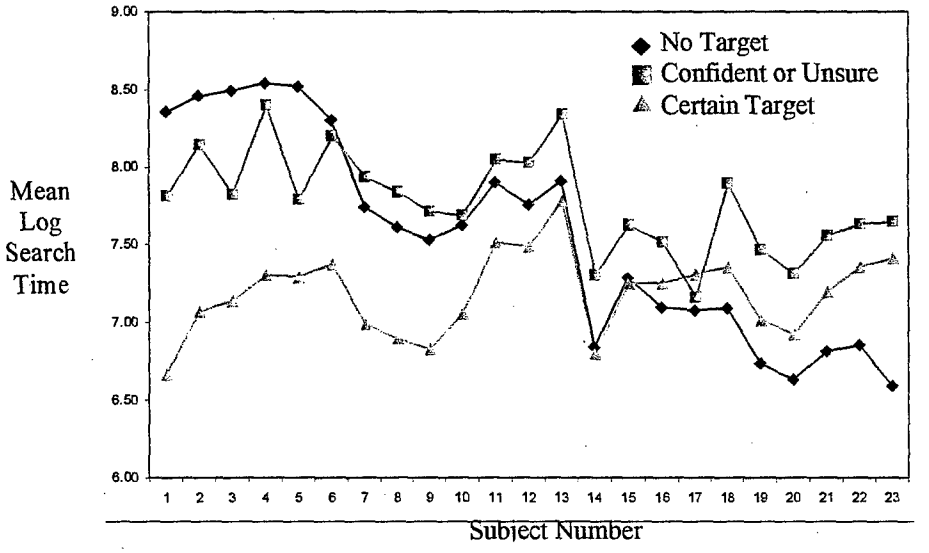

Fig. 1: Log Mean Search Time by Response Rating by Subject

Wolfe's Guided Search model [5] holds that pre-attentive processing builds up a queue of suspect locations whose activation level (based on similarity to the targets) is above the observers threshold, which are then inspected in order of decreasing activation. In the Guided Search model, quitting occurs when the list of suspect locations is exhausted. This implies that the time-to-detect will tend to be shorter than the time-to-quit. The experimental data contradicts the exhaustive search model.

The proposed model implies that the quitting and detection outcomes are in a race: the observer reports a detection or quits depending on whether he detects a target before parsing the scene into recognized background regions and objects. But it is not an independent race. Rejecting background regions makes progress towards quitting, but also guides and focuses search. Noticing suspect locations to examine contributes to quitting when they are rejected, but also contributes to the target detection outcome.

In a race process, the probability of detection for a target-in-background image can be expressed as a function of the search time for the target-in-background image and for the corresponding background-only image. Search time for a target-in-background image can be expressed as a function of the probability of detection and the search time for the corresponding background only image. The probability of detection is the probability that a random variable having a lognormal distribution with mean and standard deviation of log search time for the target-in-background image, μ_{TB} and σ_{TB}, is less than a random variable having a lognormal distribution with mean and standard deviation of log search time for the background-only image, μ_B and σ_B.

$$P_d = 2 - 1 / N\left[\frac{\mu_B - \mu_{TB}}{\alpha(\sigma_B^2 + \sigma_{TB}^2)^{1/2}}\right] \quad (1)$$

where P_d is the probability of detection for the target-in-background image, N and N^{-1} are the standard normal distribution and its inverse. The free parameter α represents the effects of correlation between the processes leading to quitting and detection. Its value, estimated from the empirical data, is 0.6. A value of unity would have indicated complete independence, whereas a value approaching zero would have indicated near perfect correlation.

The equation for P_d as a function of search time parameters can be inverted to compute the mean log search time for the target-in-background image. The standard deviation of log search time for the target-in-background image is replaced with a constant, β, equal to the average value of σ_{TB} over the set of target-in-background images.

$$\mu_{TB} = \mu_B - \alpha\,(\sigma_B^2 + \beta^2)^{1/2}\, N^{-1}\left[\frac{1}{2 - P_d}\right] \quad (2)$$

Figure 2 shows a scatter plot of the experimental estimate of the probability of detection plotted against the probability of detection predicted by the search times and the race model. The 1151 images were organized into 88 groups. Each group represents all of the variations of the 44 base images. However, variations with the unaltered, baseline targets, and the modified, low-contrast targets, were not mixed. The results show that the race model produces an accurate estimate of probability of detection.

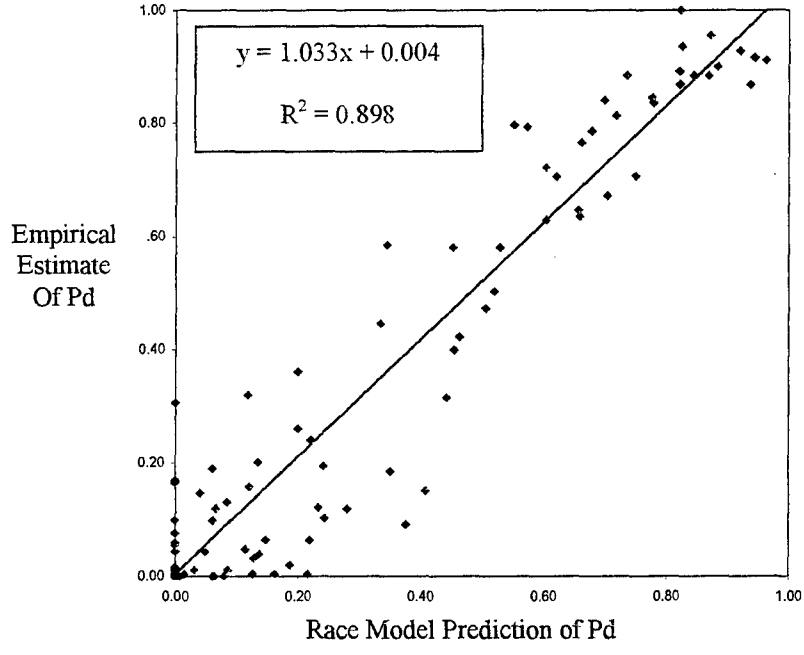

Fig. 2: Race Model Predictions of Pd vs. Empirical Estimate of Pd

Figure 3 shows the scatter plot of the experimental versus predicted mean log search time. The model predictions are highly correlated with the observed data, but the model does not provide as accurate an estimate of search time. The model prediction accounts for only 68.5% of variance in the observed mean log search times.

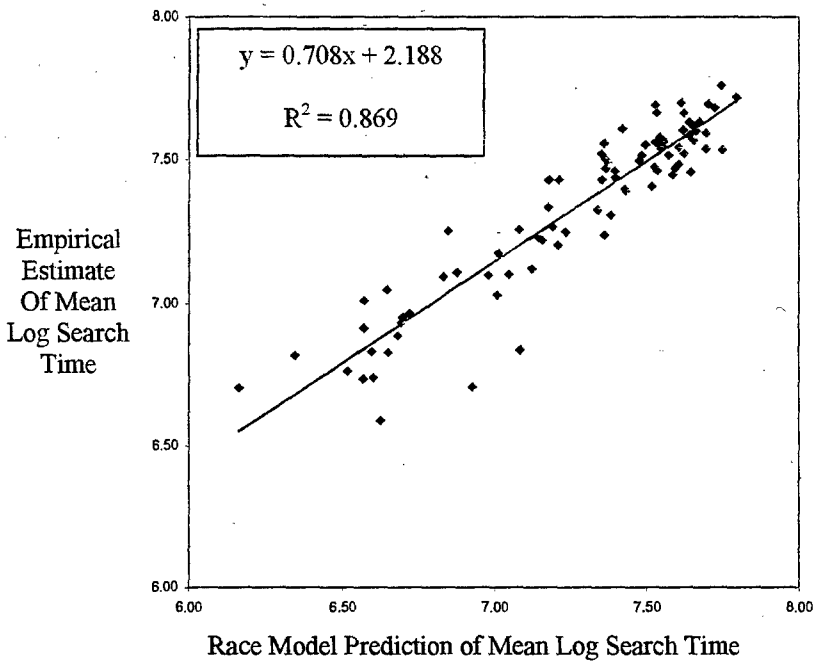

Fig. 3: Race Model Predictions vs Empirical Estimate of Mean Log Search Time

Figure 4 shows a scatter plot of the experimental mean log search time versus the empirical estimate of the probability of detection. The data show that the simple linear regression model provides as good an estimate of mean log search time as does the race model.

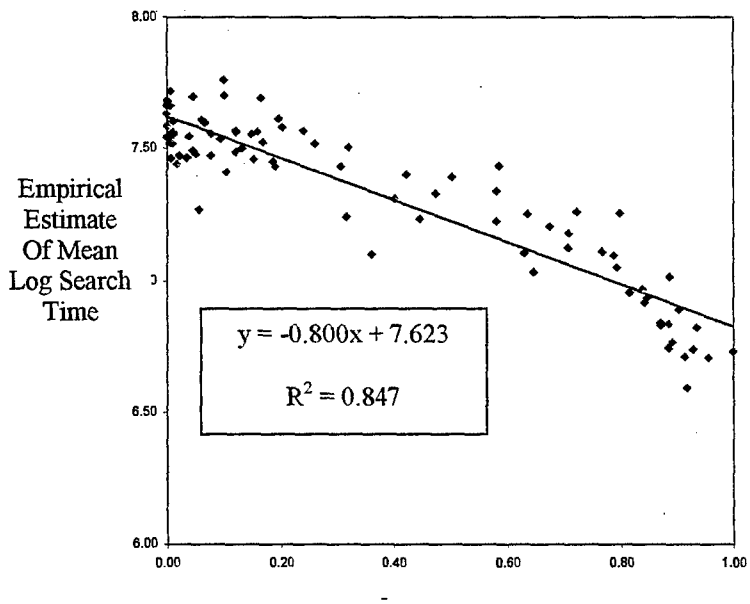

Fig. 4: Empirical Estimate Relationship between Pd and Mean Log Search Time

4 Target Signature Metric

The proposed theory of figure-ground segregation in target detection holds that pre-attentive vision organizes the target into either one region (a silhouette) or two regions (a light or illuminated region and a dark or shadowed region). Visual perception organizes the target into a dominant region and its complement. The dominant region is the region of the target with the greatest distinctiveness. The complement is the difference between the entire target and the dominant region. The perceptual organization does not depend on the combined distinctiveness of the two regions. Attentive visual processing imposes overall target bounding, and categorizes the object based on the joint contribution of the primary region and its complement.

The distinctiveness of a region is a function of the local contrast across the region boundary, the local clutter in the vicinity of the boundary (the variation within the region), and the size of the region. Local contrast and clutter are a function of the region size and shape. Region distinctiveness does not incorporate shape information value. Shape information contributes to attentive discrimination, but not pre-attentive segmentation.

The value of the signature for attentive object discrimination is a function of the combined region distinctiveness, the combined shape information magnitude, and the combined region area. It is not the sum of the value from the two regions, but the value of the combination of the two complementary regions.

The previously-developed target signature metric [1] required the user to define the canonical front, side and top regions of the vehicle. The rationale was that these regions defined the 3D structure of the vehicle, and they defined the 3D appearance of the vehicle since they had significantly different surface normal vectors with respect the source of illumination.

In practice, there were difficulties with this approach. It required a complex set of rules to deal with rounded turrets, shadows, and flat facets that were in the cardinal planes. It increased the user burden, and increased the amount of user subjectivity. Furthermore, the region organization was appropriate only for visual images. It did not apply to thermal images, which have different characteristic regions resulting from internal and external heat sources.

The revised signature metric still requires that the user designate where the object is for which he wants the metric computed. The model does not find the target outline. The target outline includes the target's shadow. In the test analysis, the target mask covered the entire projection of the vehicle onto the plane, including those parts of the scene that were foreground obscuring portions of the target. Further research is needed to address whether or not foreground obstruction should be excluded.

The revised signature metric computation begins by searching for the appropriate perceptual organization of the target. The target is organized into either (1) a single region (the target silhouette), or (2) two regions, nominally corresponding to the illuminated and shadowed regions (Two-tone representations are sufficient for 3D shape/structure perception [6]). The organization finds the sub-region of the target with the largest value of a region distinctiveness metric, and its complement.

The algorithm divides the target into light and dark regions via a threshold on the achromatic visual channel. It searches for the threshold that produces the primary target region (either above or below the threshold) with the highest region distinctiveness metric. The theory is that pre-attentive visual processing organizes the scene into distinctive shapes based on individual region distinctiveness, not on the combined effects of adjacent regions; highly distinctive regions are perceived first, and lower distinctiveness regions later. The perceptual organization consists of the primary target region (which may be the entire target projection) and its complement (which may be null).

Figure 5 illustrates the results of the perceptual organization algorithm. It shows the original image, the target mask, and the 2-region perceptual organization. Were it not for the gun tube, the target mask alone would look like a blob not a vehicle, but the 2-tone perceptual organization has the look of a vehicle.

Fig. 5: Example Image, Target Mask, and 2-Region Perceptual Organization

The processing flow to compute the region distinctiveness metric, RDM(.), is essentially the same as that used in the previous VDM2000 model [1]. There was no change in the luminance and color appearance transform from RGB to Acd (achromatic and color-opponent) coordinates.

The vehicle signature metric combines the region distinctiveness metric with a measure of the magnitude of shape information in the component regions. The measure of the shape information of a region is based on the dispersion of the region, i.e., the variance in the location of the pixels in the region (or, equivalently, the mean squared distance from the center of mass):

$$\mathrm{Dsp}(\mathcal{R}) = \Sigma_{rc \text{ in } \mathcal{R}} [\, (r - \mu_r)^2 + (c - \mu_c)^2 \,] \,/\, \mathrm{Count}(\mathcal{R}) \qquad (3)$$

The dispersion divided by the area is a measure of the compactness of the region. A commonly used alternative measure of the shape complexity or compactness of a region is the perimeter squared divided by the area. However this simple metric has some serious shortcomings: a shape that is almost a circle but with ragged edges can have a very large value, and the metric for a shape consisting of disjoint regions is independent of the configuration of the disjoint regions. The dispersion metric does not have these problems.

The vehicle signature metric is equal to the sum of the region distinctiveness metrics, times the sum of the region dispersions, Dsp(.), divided by the sum of the region areas, Area(.) for the two perceptual regions $\mathcal{R}1$ and $\mathcal{R}2$. The metric for the vehicle is based on the combination of the complementary regions:

$$M = (\mathrm{RDM}(\mathcal{R}1) + \mathrm{RDM}(\mathcal{R}2)) * (\mathrm{Dsp}(\mathcal{R}1) + \mathrm{Dsp}(\mathcal{R}2)) / (\mathrm{Area}(\mathcal{R}1) + \mathrm{Area}(\mathcal{R}2)) \qquad (4)$$

The psychometric function is unchanged from the earlier model formulation. The entire model has only one free parameter: the vehicle metric value for 50% probability of detection.

Figure 6 shows a scatter plot comparing the model prediction of Pd with the empirical estimate of Pd for all 800 images with targets in the cued detection experiment. Figure 7 compares Pd in the cued detection experiment with Pd from the search experiment. Figure 8 compares the model prediction of Pd in search with Pd from the search experiment. The model accounts for 72% of the variance in Pd in the cued detection experiment. Pd in the cued detection experiment accounts for 81% of the variance in Pd in the search experiment. The chain rule predicts that the model should account for 58% of the variance in Pd in the search experiment. In fact, the model accounts for 61% of the variance in Pd in the search experiment.

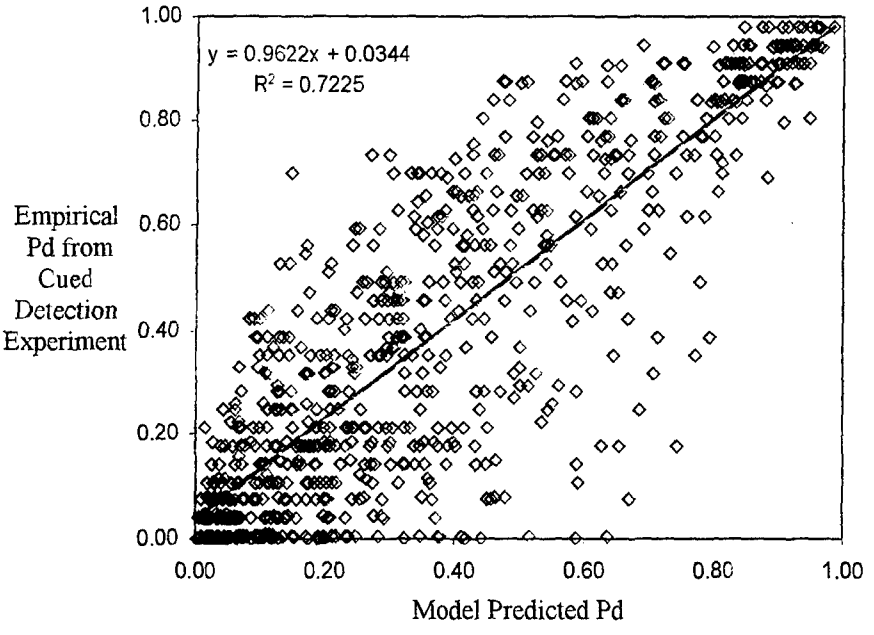

Fig. 6: Comparison of Model and Empirical Pd for Cued Detection Experiment

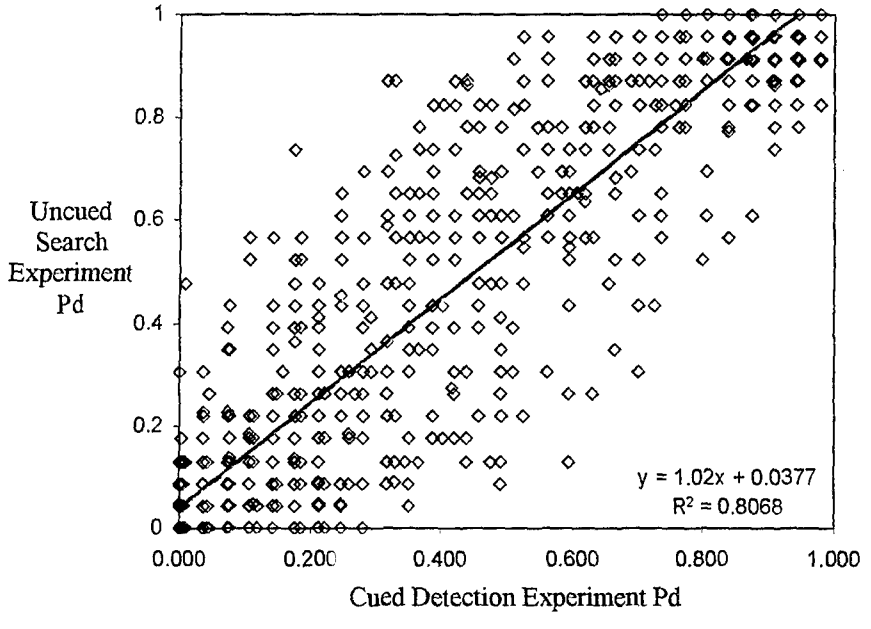

Fig. 7: Comparison of Empirical Pd for Cued Detection and Uncued Search Experiments

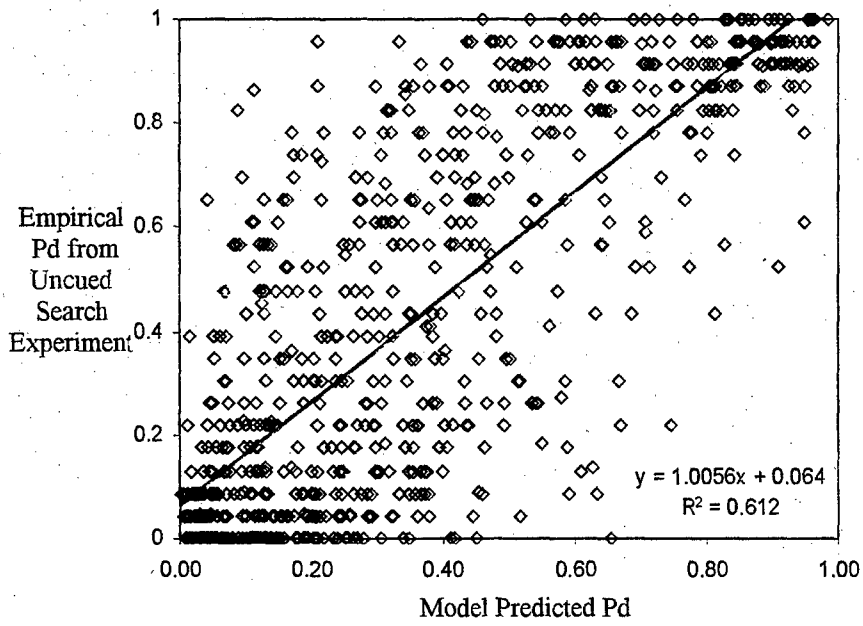

Fig. 8: Comparison of Model and Empirical Pd for Uncued Search Experiment

5 Discussion

The proposed model is derived from a theory of search behavior assembled from empirical and theoretical research in human vision and decision making, and is substantiated by comparison to the results of large-sample search and target detection experiments.

The proposed model is similar in structure, but differs in specifics, from the widely-used U. S. Army NVESD search model [7]. Both models predict the probability of detection within some specified time as the product of the probability of detection and the probability of search completion within the specified time. The proposed model uses a lognormal formulation for the distribution of search time, whereas the NVESD model uses a negative exponential formulation. The proposed model uses a constant for one parameter of the lognormal distribution, and calculates the other parameter as a linear function of the probability of detection. The NVESD model calculates the rate parameter of the negative exponential distribution as proportional to the probability of detection. The proposed model uses a simple, one parameter psychometric function. The NVESD psychometric function uses three parameters. The greatest difference between the two models is in the vehicle signature metric. The NVESD metric is a function of the target area, target-to-background contrast and range. The proposed metric is an image-based computational vision formulation that accounts for target size, shape, perceptual organization, local clutter, local color and luminance contrast across the target boundary.

The vehicle signature metric in the earlier VDM2000 model was not applicable to thermal images since it employed a simple 3D geometry appropriate to solar illumination and reflected signature. It was not applicable to long-wave infra-red images, in which the perceptual regions correspond to hot and cold physical components. The revised vehicle signature metric discovers the appropriate perceptual organization. It is potentially applicable to thermal images, but has not been tested.

The refined signature metric still has one important limitation. The problem occurs when an edge of the target is in line with a linear feature in the scene, and when the contrast across that target edge is close to the contrast across the background feature. When this occurs, human visual perception tends to see the target edge as a continuation of the background feature. When the contrast difference is low, the perceptual organization that segments the background feature dominates the segmentation of the target, and probability of detection is low. The computational method has no way to detect this condition, and so

overestimates the probability of detection. This condition occurred in approximately 100 of the 800 images used in the perception experiments.

Acknowledgments

This research was funded by the U.S. Army TACOM under Small Business Innovation Research contract DAAE07-97-C-X101.

References

1. Witus, G., Karlsen, R. and Gerhart, G. 2001. The VDM2000 visual vehicle detection model – theory and validation. *Proceedings of the 12th Annual Ground Target Modeling and Validation Conference*, Houghton, MI. 260-268.

2. Cartier, J. S. and Hsu, D. H. 1995. Human visual search: a two state process. *Proc. SPIE Vol. 2470*, p. 58-68, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing VI, Gerald C. Holst; Ed.

3. Ulrich, R. and Miller, J. 1993. Information processing models generating lognormally distributed response times. *Journal of Mathematical Psychology*. 37, 513-525.

4. Shapiro, S. S., and Wilk, M. B. 1965. An analysis of variance test for normality. *Biometrika*. 52, 591.

5. Wolfe, J. M. 1994. Guided Search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review*. 1, 202-238.

6. Moore, C., and Cavanagh, P. 1998. Recovery of 3D volume from 2-tone images of novel objects. *Cognition*. 67:45-71.

7. O'Kane, B. L. 1995. Validation of prediction models for target Acquisition with electro-optical sensors. In *Vision Models for Target Detection and Recognition*, E. Peli, Ed. World Scientific Press. 192-218.